PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
G01N 27/00, 27/26, 33/543, 33/551, C12Q 1/68, A01N 1/02, C12M 1/00

A1

(11) International Publication Number: WO 98/08083

(43) International Publication Date: 26 February 1998 (26.02.98)

(21) International Application Number: PCT/US97/14372

(22) International Filing Date: 13 August 1997 (13.08.97)

(30) Priority Data:
08/700,182 20 August 1996 (20.08.96) US

(71) Applicant (*for all designated States except US*): MOTOROLA INC. [US/US]; 1303 East Algonquin Road, Schaumburg, IL 60196 (US).

(72) Inventors; and
(75) Inventors/Applicants (*for US only*): REBER, William, L. [US/US]; 1029 Bucaneer Road #6, Schaumburg, IL 60916 (US). ACKLEY, Donald, E. [US/US]; 2033 Cambridge Avenue, Cardiff, CA 92007 (US). PERTTUNEN, Cary, D. [US/US]; 11764 Raintree Court, Shelby Township, MI 48315 (US).

(74) Agents: SARLI, Anthony, J., Jr. et al.; Motorola Inc., Intellectual Property Dept., 1303 East Algonquin Road, Schaumburg, IL 60196 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, HU, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) **Title:** METHOD AND APPARATUS FOR DETECTING PREDETERMINED MOLECULAR STRUCTURES IN A SAMPLE

(57) **Abstract**

A predetermined molecular structure in a sample is detected by sensing a pattern in which the sample binds to an array of binding sites in a molecular detection device (20), and comparing the pattern to a reference pattern to detect the predetermined molecular structure in the sample (22). In one embodiment, the reference pattern is generated by sensing a pattern in which a reference sample containing the predetermined molecular structure binds to a like array of binding sites. In another embodiment, the reference pattern is generated by predicting a pattern in which the predetermined molecular structure binds to the array of binding sites.

METHOD AND APPARATUS FOR DETECTING PREDETERMINED MOLECULAR STRUCTURES IN A SAMPLE

Field of the Invention

The present invention relates to methods and system for molecular detection.

Background of the Invention

An increased effort has been directed toward the development of chips for molecular detection. Typically, a molecular detection chip includes a substrate on which an array of binding sites is arranged. Each binding site, or hybridization site, has a respective molecular receptor which binds or hybridizes with a molecule having a predetermined structure.

A sample solution is applied to the molecular detection chip, and molecules in the sample bind or hybridize at one or more of the binding sites. The particular binding sites at which hybridization occurs are detected, and one or more molecular structures within the sample are subsequently deduced.

Of great interest are molecular detection chips for gene sequencing. These chips, often referred to as DNA chips, utilize an array of selective binding sites each having respective single-stranded DNA probes. A sample of single-stranded DNA fragments, referred to as target DNA, is applied to the DNA chip. The DNA fragments attach to one or more of the DNA probes by a hybridization process. By detecting which DNA probes have a DNA fragment hybridized thereto, a

FIG. 6 illustrates a reference pattern for detecting an a-c-t nucleotide sequence in a sample;

FIG. 7 is an example of a pattern generated by a sample having an unknown molecular structure;

FIG. 8 is a flow chart of additional steps which can be utilized to detect the predetermined molecular structure;

FIG. 9 is a block diagram of an apparatus for detecting a predetermined molecular structure in a sample; and

FIG. 10 is a flow chart of an embodiment of a method of gene discovery in accordance with the present invention.

Detailed Description of a Preferred Embodiment

Embodiments of the present invention advantageously provide improved information processing approaches to detecting predetermined molecular structures using a miniaturized device having an array of biological sensors. Just as semiconductor devices are designed to perform specific functions, a diagnostic device in accordance with the present invention is designed to perform one or more specific diagnostic tests.

FIG. 1 is a flow chart of an embodiment of a method of detecting a predetermined molecular structure in a sample. In general, the method can be utilized for detection of a variety of molecular structures in a variety of different types of samples. Examples of the different types of samples include, but are not limited to, medical samples, environmental samples, agricultural samples, and other samples applicable to diagnostics.

hybridizes with a molecule having a predetermined structure.

Each molecular receptor typically includes a biological or synthetic molecule having a specific affinity to the molecule to be detected. Of particular interest is a molecular receptor having a chain of at least one nucleotide to hybridize with a molecule having a complementary chain of at least one nucleotide. Here, for example, the molecular receptor can include a DNA probe for detecting a corresponding, complementary DNA sequence in the sample.

It is noted, however, that the scope of the invention is not limited to sensing the hybridization of DNA molecules. For example, embodiments of the present invention can be utilized to detect RNA hybridization and antibody-antigen binding events. As another alternative, the molecular detection device can include an array of detection sites, such as in the context of an oligonucleotide ligation assay (OLA). Using a ligase chain reaction, pairs of oligonucleotides are utilized to amplify a selected oligonucleotide sequence. To detect the selected oligonucleotide sequence, a corresponding detection site is screened for full-length ligated oligonucleotides using any of the sensing approaches described herein.

As indicated by block 12, an optional step of tagging molecules within the sample is performed. Each molecule is tagged with a member which can be sensed by the molecular detection device. Such members are commonly referred to in the art as tags, markers, and labels. Examples of such members include, but are not limited to,

After hybridization, an optional step of removing unwanted molecules from the binding sites can be performed, as indicated by block 18. The step of removing unwanted molecules can be performed by generating an electric field having the same polarity as the charge of the unwanted molecules. The electric field acts to repel unwanted molecules from the binding sites. As an alternative to, or in conjunction with, the field-based approach, a thermally-assisted approach can be utilized to remove unwanted molecules. Here, the temperature at the binding sites is raised, in dependence upon a melting temperature, to dissociate partially-bound molecules from the molecular receptors. Regardless of the approach utilized, the unwanted molecules to be dehybridized can include unbound molecules and partially-bound (i.e. non-specifically bound) molecules.

Typically, the step of removing unwanted molecules does not remove all unwanted molecules from the binding sites. This step is beneficial, however, in improving the accuracy of detection as outlined in subsequent steps.

As indicated by block 20, the method includes a step of sensing a pattern in which the sample binds to an array of binding sites in a molecular detection device. The pattern can be sensed using a variety of approaches, including but not limited to, optical approaches, radioactive-sensing approaches, electronic approaches, and magnetic approaches. The specific approach utilized depends upon the type of tagging member attached to the molecules in the sample.

Preferably, the step of sensing the pattern includes sensing an intensity or a magnitude of

determining a difference between the pattern and the reference pattern. Here, the predetermined molecular structure can be detected when a measure of the difference is within a predetermined range.

For optical sensing embodiments, the step of comparing includes a step of comparing at least one image of the pattern to an image of the reference pattern.

As indicated by block 24, the method optionally comprises a step of determining a confidence level of detecting the predetermined molecular structure in the sample. The confidence level indicates a degree of significance of the result obtained in the step of comparing in block 22.

To screen the sample for a plurality of different molecular structures, the steps indicated by blocks 22 and 24 can be repeated for a plurality of different reference patterns. Here, for example, a genomic sample can be screened to determine if it contains any of a plurality of predetermined base sequences.

FIG. 2 is a flow chart of an embodiment of a method of generating the reference pattern. Typically, the reference pattern is generated prior to performing the steps indicated in FIG. 1.

As indicated by block 30, the method includes a step of providing a reference device having a like array of binding sites as the molecular detection device used for detection. If desired, the same molecular detection device can be utilized for generating the reference pattern and for subsequent detection of an unknown molecular structure in a sample. Typically, however, another like device is utilized.

molecular structure. Alternatively, the steps indicated by blocks 30, 32, 34, 36, 38, 40, and 42 can be repeated to apply the same predetermined molecular structure to a number of like reference devices. Either approach may be utilized to provide a plurality of reference patterns for the same predetermined molecular structure.

Sensed patterns formed by a sample having an unknown molecular structure can be compared to each of the above-described plurality of reference patterns, or a statistical model thereof, to detect the predetermined molecular structure in the sample.

FIG. 3 is a flow chart of another embodiment of a method of generating the reference pattern. As indicated by block 50, the method includes a step of determining an architecture of the array of binding sites of the molecular detection device. This step can include determining a layout of the binding sites, and determining the type of molecular receptor at each of the binding sites.

As indicated by block 52, the method includes a step of predicting a reference pattern in which the predetermined molecular structure binds to the array of binding sites. The reference pattern is predicted based upon the predetermined molecular structure and the architecture of the molecular detection device. Preferably, the reference pattern includes a predicted intensity of binding at each of the binding sites.

Regardless of the approach taken, the reference pattern acts as a novelty filter which is predictive of a successful or a desirable test result.

as binding sites indicated by reference numeral 64, have a lesser intensity of binding. Those binding sites having two mismatching complementary bases, such as those indicated by reference numeral 66, have an even lesser intensity of binding. Binding sites with no matching complementary bases, such as those indicated by reference numeral 68, have a low intensity of binding.

FIG. 6 illustrates a reference pattern for detecting an a-c-t nucleotide sequence in a sample. The reference patterns in FIGS. 5 and 6 can be sensed using a reference sample, or can be predicted based on the number of mismatching bases at each binding site.

For purposes of illustration, the sequences in this example are assumed to have a specific orientation. As a result, an a-c-t sequence and a t-c-a sequence do not specifically hybridize at the same binding site. It is noted, however, that this should not be construed as a limitation in the scope of the present invention.

FIG. 7 is an example of a pattern generated by a sample having an unknown molecular structure. The pattern is generated by applying a sample of tagged single-stranded DNA molecules to the molecular detection device, allowing the molecules to hybridize to the binding sites, and optionally removing unwanted molecules.

The resulting pattern shows a high intensity of binding at a t-g-a site 70. If standard detection techniques were utilized, one would conclude that the sample includes an a-c-t nucleotide sequence (i.e. the complement of t-g-a). However, the overall pattern is better correlated to the reference pattern for the a-c-g nucleotide

selected binding sites, or can be modified for all of the binding sites.

Thereafter, a step of sensing a pattern, indicated by block 82, is performed. By repeatedly raising the temperature and sensing a resulting pattern, a plurality of temperature-dependent patterns is generated.

As indicated by block 84, a step of comparing at least one of the plurality of temperature-dependent patterns to a corresponding at least one of a plurality of temperature-dependent reference patterns is performed. Here, for example, each of the temperature-dependent patterns can be compared to a corresponding one of the temperature-dependent reference patterns. Alternatively, only selected ones of the temperature-dependent patterns can be compared to corresponding reference patterns. A correlation measure and/or a difference measure is computed based on this comparison. A predetermined molecular structure is detected when the measure is within a predetermined range.

In one embodiment, the temperature-dependent pattern having a greatest variability of intensity is selected for comparison. The variability of intensity is greatest at a temperature which dissociates many non-specifically-bound molecules, but does not significantly dissociate specifically-bound molecules. This pattern can be compared with a corresponding reference pattern to detect a predetermined molecular structure. It is noted that a variety of different measures of variability can be utilized, including but not limited to, sample variance and sample standard deviation.

FIG. 9 is a block diagram of an apparatus for detecting a predetermined molecular structure in a

present invention. As indicated by block 100, the method includes a step of sensing a pattern of detection for a sample applied to a molecular detection device having a plurality of detection sites. The sample is taken from a first species having unknown genes to be discovered. Any of the various approaches described herein can be utilized for sensing the pattern.

As indicated by block 12, the method includes a step of determining an architecture of the plurality of detection sites of the molecular detection device.

A step of reading a nucleotide sequence from a database is performed, as indicated by block 104. In general, any nucleotide sequence can be read. Of particular interest, however, is a nucleotide sequence from a second species other than the first species. For example, the nucleotide sequence can include a gene from a fruit fly, while the sample in which gene discovery is to be performed is from a human.

As indicated by block 106, the method includes a step of predicting a reference pattern which would be detected if the nucleotide sequence were applied to the molecular detection device. As indicated by block 108, a step of comparing the pattern to the reference pattern is performed to determine whether the nucleotide sequence is within the sample.

The steps of reading a nucleotide sequence from the database, predicting a reference pattern for the nucleotide sequence, and comparing the pattern to the reference pattern are repeated to discover the presence of genes across different species.

Claims

1. A method of detecting a predetermined molecular structure in a sample, the method comprising the steps of:

sensing a pattern in which the sample binds to an array of binding sites in a molecular detection device; and

comparing the pattern to a reference pattern to detect the predetermined molecular structure in the sample.

2. The method of claim 1 further comprising the step of determining a confidence level of detecting the predetermined molecular structure in the sample.

3. The method of claim 1 wherein the step of comparing includes at least one of determining a difference between the pattern and the reference pattern, determining a correlation between the pattern and the reference pattern, comparing at least one image of the pattern to an image of the reference pattern, comparing a plurality of temperature-dependent patterns to a plurality of temperature-dependent reference patterns.

4. The method of claim 1 wherein an intensity level of at least one of the binding sites is indicative of at least one of molecules bound at a respective binding site, a binding strength at a respective binding site, and a melting temperature at a respective binding site.

array of binding sites or by predicting a pattern in which the predetermined molecular structure binds to the array of binding sites.

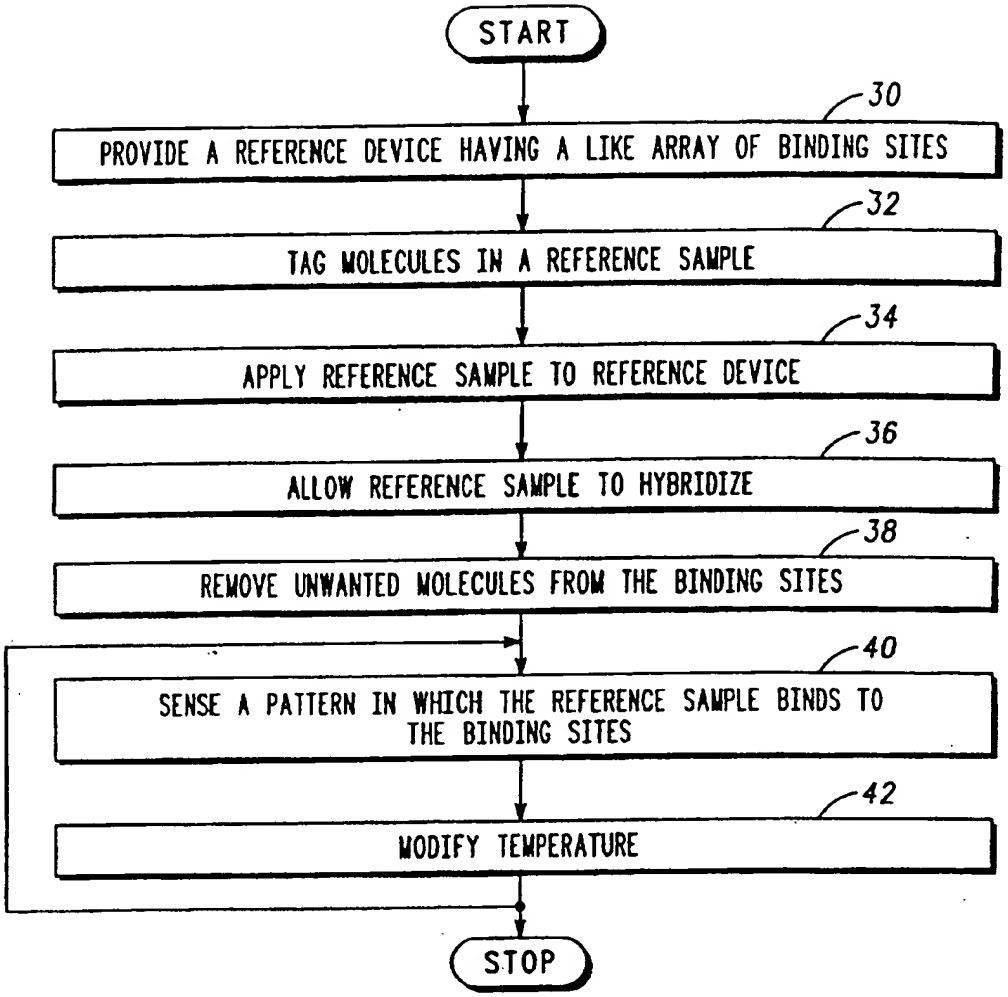
START
30
PROVIDE A REFERENCE DEVICE HAVING A LIKE ARRAY OF BINDING SITES
32
TAG MOLECULES IN A REFERENCE SAMPLE
34
APPLY REFERENCE SAMPLE TO REFERENCE DEVICE
36
ALLOW REFERENCE SAMPLE TO HYBRIDIZE
38
REMOVE UNWANTED MOLECULES FROM THE BINDING SITES
40
SENSE A PATTERN IN WHICH THE REFERENCE SAMPLE BINDS TO THE BINDING SITES
42
MODIFY TEMPERATURE
STOP

FIG.2

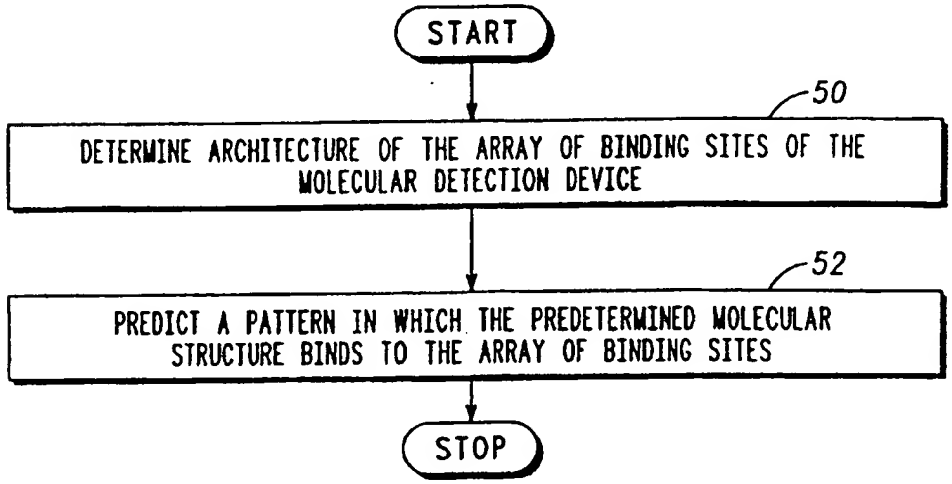
START
50
DETERMINE ARCHITECTURE OF THE ARRAY OF BINDING SITES OF THE MOLECULAR DETECTION DEVICE
52
PREDICT A PATTERN IN WHICH THE PREDETERMINED MOLECULAR STRUCTURE BINDS TO THE ARRAY OF BINDING SITES
STOP

FIG.3

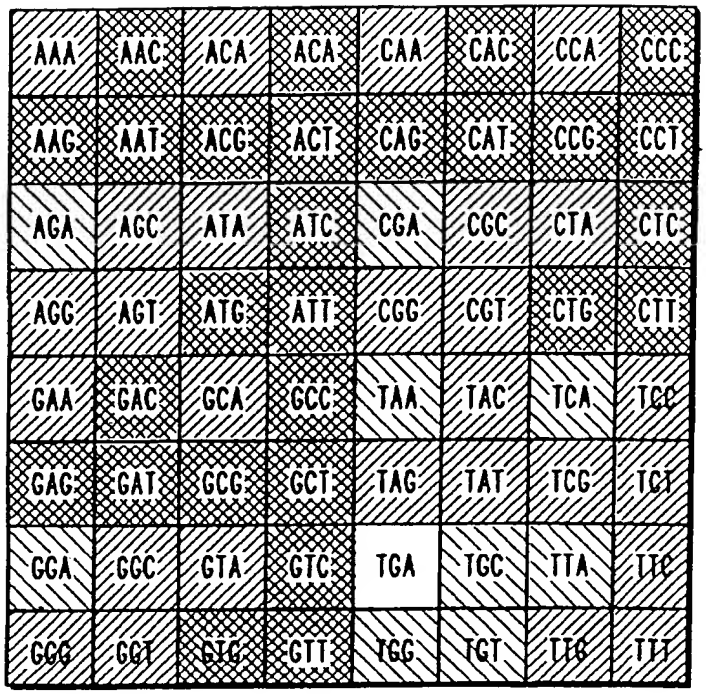
AAA AAC ACA ACA CAA CAC CCA CCC
AAG AAT ACG ACT CAG CAT CCG CCT
AGA AGC ATA ATC CGA CGC CTA CTC
AGG AGT ATG ATT CGG CGT CTG CTT
GAA GAC GCA GCC TAA TAC TCA TCC
GAG GAT GCG GCT TAG TAT TCG TCT
GGA GGC GTA GTC TGA TGC TTA TTC
GGG GGT GTG GTT TGG TGT TTG TTT

FIG.6

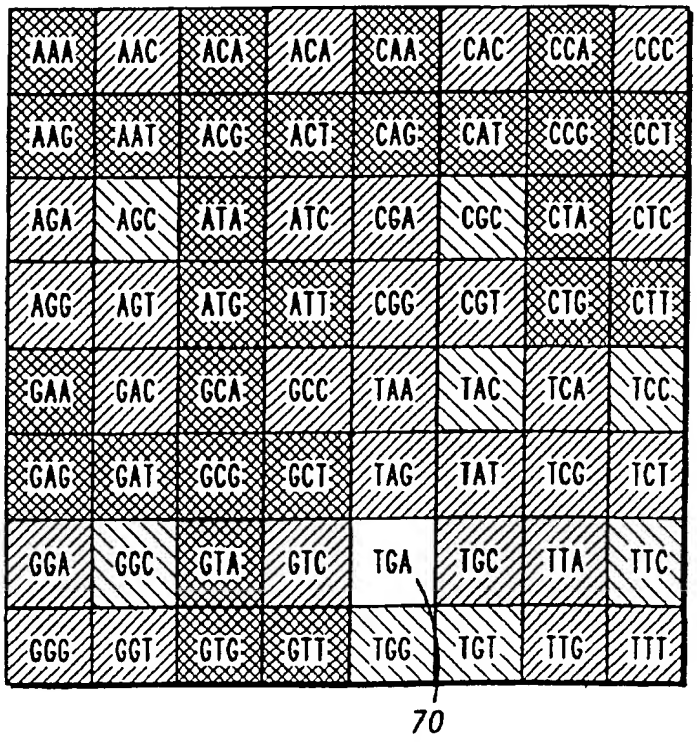
AAA AAC ACA ACA CAA CAC CCA CCC
AAG AAT ACG ACT CAG CAT CCG CCT
AGA AGC ATA ATC CGA CGC CTA CTC
AGG AGT ATG ATT CGG CGT CTG CTT
GAA GAC GCA GCC TAA TAC TCA TCC
GAG GAT GCG GCT TAG TAT TCG TCT
GGA GGC GTA GTC TGA TGC TTA TTC
GGG GGT GTG GTT TGG TGT TTG TTT
70

FIG.7

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US97/14372

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) :Please See Extra Sheet.
US CL :Please See Extra Sheet.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 204/153.1, 400, 403; 435/6, 283.1, 287.1, 287.2; 436/518, 524, 807, 809

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

APS, CAPLUS, MEDLINE, WPIDS, SCISEARCH, MEDLINE, EMBASE, BIOSIS

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 5,532,128 A (EGGERS et al) 02 July 1996. See entire document.	1-10
X	PEASE AC. Light-generated Oligonucleotide Arrays for Rapid DNA Sequence Analysis. Proc. Natl. Acad. Sci. May 1994. Vol 91. pages 5022-5026, especially Abstract.	1-10
X	STIMPSON DI. Real-time Detection of DNA Hybridization and Melting On Oligonucleotide Arrays by using Optical Wave Guides. Proc. Natl. Acad. Sci. July 1995. Vol 92. pages 6379-6383, especially Abstract	1-10
A,P	US 5,605,662 A (HELLER et al) 25 February 1997.	1-10

[X] Further documents are listed in the continuation of Box C. [ ] See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E"	earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search 24 NOVEMBER 1997	Date of mailing of the international search report 23 DEC 1997
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer VERY [illegible] Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet)(July 1992)*

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US97/14372

A. CLASSIFICATION OF SUBJECT MATTER:
IPC (6):

G01N 27/00, 27/26, 33/543, 33/551; C12Q 1/68; A01N 1/02; C12M 1/00;

A. CLASSIFICATION OF SUBJECT MATTER:
US CL :

204/153.1, 400, 403; 435/6, 283.1, 287.1, 287.2; 436/518, 524, 807, 809

Form PCT/ISA/210 (extra sheet)(July 1992)*